The questions below should be carried out in Python or R. The code should be returned within 5 full business days after the exercise is received. Submissions are evaluated based on correctness, the structure and clarity of the code, as well as the style of the output and data visualization.

I. Optimal Investing

a). Download stock price data for 10+ equities from free online sources.

b). Use these stocks to compute a mean-variance efficient frontier. How did you compute the expected return for each stock? The covariance matrix? What start/end date did you use for the return series? What frequency are the returns? Visualize the covariance matrix. Report all the expected returns and covariances as annualized quantities.

c). Explain what the efficient frontier means. Which portfolio would you invest in and why?

d). Add short-sale constraints and plot the constrained efficient frontier together with the one from b). Explain any differences you see.

e). If you wanted to invest in 200 equities, explain any difficulties in computing the covariance matrix and how these can be overcome.

II. Linear Regression

Load info.txt and performance.csv into R. The file performance.csv has columns ID, Date, and Performance, which are monthly returns for each instrument identified by ID and each Date since January 1, 1990. The file info.txt has static information about each instrument identified by ID, including Name, Currency, and (Geographic) Focus.

a). Using the R package sqldf, produce a data frame with Date, Performance from performance.csv and corresponding Name from info.txt.

b). Using the package openxlsx, write the data frame in b) to an Excel file.

c). Regress Stock CZ against Index, and interpret the results. Are the coefficients significantly different from zero? What is the beta?

d). Plot the Stock CZ returns against the fitted returns from the linear regression in d). Use plotly.

III. Correlated Contribution Times

Say you invest in a PE fund that will invest your commitment equally in 10 companies. Ignore management fees. Let $\tau_1, \ldots, \tau_{10}$ denote the time when each company is purchased.

a). Assume $\tau_i \sim_{iid} \exp(\lambda)$ for $\lambda = 1/2.5$. Simulate the contributions for one fund. Plot the cumulative contributions. What is the interpretation of λ?

b). Now simulate 100 funds and compute the average cumulative contribution curve. Plot it. What is this curve?

c). In practice, contribution times may be dependent. If I find one excellent company to buy, it's more likely that I will soon find another excellent company to purchase. Find a model that will allow you to simulate dependent contribution times. Describe the model and the intuition behind the dependence. How did you set the parameters? Simulate one fund and plot it.

d). Simulate 100 funds and compute the average cumulative contribution curve. How does it differ from b)?

e). For the models in a) and c), compute the maximum contribution in any one calendar year across many simulations. Plot the distribution of the maximum. LPs need to hold cash in order to meet these contributions. Under which model does the client need to hold less cash?